

修 士 論 文 の 和 文 要 旨

大学院 電気通信学 研究科 博士前期課程 システム工学 専攻		
氏 名	秋山 哲史	学籍番号 0535002
論 文 題 目	ベクトル空間モデルに基づく単語の意味表現の性質と構築手法	
<p>要 旨</p> <p>様々な言語処理技術において、単語の意味をどのように扱うかは重要な問題である。単語の意味を表現する手法には様々なものがあり、その一つとして、ベクトル空間モデルに基づいた手法がある。これは、テキストデータにおける単語の出現頻度等を基に単語を多次元空間上に配置する手法であり、その多次元空間上での近さが単語同士の関係の強さを反映する。</p> <p>ベクトル空間モデルに基づいた意味表現の構築手法は、今日様々なものが提案されており、異なる手法で構築された意味表現の間には、当然ながら性質の差異が予想されるが、意味表現の性質の比較や、構築手法と性質の関係についての研究はほとんど行われていない。その中で、鈴木は語義的類似と、連想関係から語義的類似に相当するものを省いた連想的類似の二つの類似性を定義することで意味表現の性質の比較を行っている。しかし、連想関係には、反意語、類義語関係といった語義的類似の他にも、統語関係や熟語関係などの文脈上での近さや、「犬」と「動物」のような概念的な上下関係など、単語間の様々な関係が包含されている。意味表現の性質を明らかにする上では、それらの関係についても区別して扱うべきである。</p> <p>本研究では、異なる手法を用いて構築された様々な意味表現の性質の差異を明らかにするとともに、手法と性質との関係を明らかにすることを目的とし、単語間の様々な関係において、異なる手法で構築した意味表現の比較実験を行った。テキストデータとして国語辞典および新聞記事を用い、出現頻度方式 (TF)、共起頻度方式 (CO)、概念ベース方 (CB)、Sliding Window方式 (SW) の4方式を用いて構成した特徴行列および、これらを、シソーラス圧縮 (TR)、特異値分解 (SVD)、Random Indexing (RI) の3手法を用いて次元圧縮して構築した意味表現を比較対象とした。また、大規模な連想実験の結果データをアンケートにより様々な関係に分類し、それぞれの関係ごとに各意味表現の比較を行った。</p> <p>比較実験の結果、「近接的關係」は新聞記事・TF、「上位-下位關係」は国語辞典・CO・TR、「語義的關係」は国語辞典・TFの組合せがそれぞれ最もよく表現できることがわかった。</p> <p>またさらに、意味表現の構築にWebテキストを利用することの有効性を検証し、WEBテキストにTF手法を用いたとき、新聞記事に対して同手法を用いた場合と同程度の性能が得られることを示した。</p>		